Dell OpenManage Power Center's Power Policies for 12th-Generation Servers

This Dell white paper describes the advantages of using the Dell OpenManage Power Center to set power policies in a data center.

Author(s)

Hai Phung

Ross Burns

Paula Desmond

This document is for informational purposes only and may contain typographical errors and technical inaccuracies. The content is provided as is, without express or implied warranties of any kind.

© 2012 Dell Inc. All rights reserved. Dell and its affiliates cannot be responsible for errors or omissions in typography or photography. Dell, the Dell logo, and PowerEdge are trademarks of Dell Inc. Intel and Xeon are registered trademarks of Intel Corporation in the U.S. and other countries. Microsoft, Windows, and Windows Server are either trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell disclaims proprietary interest in the marks and names of others.

August 2012| Rev 1.0

Contents

Executive Summary 5
Introduction 5
Dell 12th-Generation Servers Employ Intel Node Manager for Power Monitoring and Management 5
Power Capping Based on Real-Time Power Measurement 5
Setting an Efficient Capping Value 6
Power Capping a Group of Servers 8
Power Center can Aggregate the Data and Set the Power Capping at any Group Level 8
Static Capping 8
Dynamic Capping 9
How to Set Power Policies using Power Center 11
Setting Static Policies 11
Setting Dynamic Policies 13
Managing Power Policies 15
Power Provisioning Using Real-Time Versus Nameplate Data 15
Using Power Policies for Emergency Management 16
Setting a Low or Untenable Power Cap on the Server 17
Peak Power Reporting and Power Capping 18
Using iDRAC in Conjunction with Power Center 19
Power Capping Overlap and Logical Groups 19
Conclusion 19
References 19

Tables

Table 1. The power is dynamically adjusted among servers in a group 10

Figures

Figure 1. Power operation range and capping value 7
Figure 2. Power Center - Power operation range and capping value 7
Figure 3. The power data is aggregated for group capping 8
Figure 4. Static capping in Power Center 9
Figure 5. Setting the static policy type 12
Figure 6. Setting group capping value in static policy 13
Figure 7. Setting priorities in dynamic policy 14
Figure 8. View, manage, and mitigate all power policies from one place 15
Figure 9. Power consumption is reduced due to EPR 17

Figure 10. Maximum power consumption spikes above the capping value 18

Executive Summary

Dell OpenManage Power Center uses Intel Node Manager for real-time power monitoring and management on Dell 12th-generation servers. IT administrators can set power capping values for individual servers as well as for groups of servers in order to manage power consumption within their environment. Power Center supports both static and dynamic power capping for a group modeled on a physical infrastructure (data center, room, aisle, or rack) as well as logical-based groups. Additionally, power policies can be layered and scheduled at various times, with auto-calculation and application of the most restrictive power policies to a group or device.

Introduction

Monitoring and controlling power consumption has become more and more important in today's data centers. Managing the power in a data center can now significantly lower energy bills and may increase server uptime. Power Center offers administrators several ways of monitoring and managing power consumption to suit their unique environments. Some specific features of Power Center include real-time power usage statistics from PMBus-enabled 11th- and 12th-generation servers, power capping on 12th-generation servers on an individual and/or group level, dynamic power allocation on groups of 12th-generation servers, and an emergency power reduction feature to immediately cap a server's power at its minimum running power.

Dell 12th-Generation Servers Employ Intel Node Manager for Power Monitoring and Management

New to 12th-generation servers, Intel's Node Manager enables the tracking of power statistics and more importantly, allows for almost real-time responses to power reduction and increased power allocation requests. Node Manager has its own processor and firmware on the server, outside of any operating system running on the system, so its performance is not affected by other running applications, and vice-versa. Node Manager, along with Dell-developed IP, enables regulation of server power consumption through the modulation of processor and memory performance through various throttling mechanisms, also reducing the power usage of other server subsystem components (memory, fan, etc.).

Power Capping Based on Real-Time Power Measurement

In Power Center, an administrator can define static power policies by setting static power caps (or a static power budget), and these caps will then be sent to Node Manager. These power caps do not change based on the power consumption of the server (or any other servers around it).

For dynamic power caps, Power Center may frequently send new power cap values to Node Manager, based on the server's power consumption and the power allocated to the server or the group(s) it is a part of. Power is monitored frequently at 1-, 3-, or 6-minute intervals, configurable by the user. In each polling interval, the minimum, maximum, and average power consumption values are retrieved from Node Manager and fed into the dynamic power policy calculations for the server. Utilizing this power data, Power Center can actively redistribute the power budget among servers to maintain the power usage within the individual and/or group dynamic power policies. When the power usage exceeds the cap for a server, Power Center uses Node Manager to throttle the power consumption, keeping it under the capping value. The throttling in Node Manager is done in a matter of milliseconds.

Setting an Efficient Capping Value

It is important to set an efficient power cap value for each server. While setting the value too high is inefficient, setting the cap value too low will impact performance and prevent the server from carrying out its computational workload effectively. Having the real-time power usage data from Power Center is a good starting point for setting an efficient capping value on the server. Power Center also displays the server's operational power range based on data leveraged from the iDRAC power management interface. During the server's power-on self-test (POST), the iDRAC power management function will determine the highest possible power consumption of a server by executing a measurement test of power consumption under a simulated maximum load, resulting in the upper power limit of the operation of the server as seen in Power Center. Power Center also collects measurement data over time, specifically the minimum power usage for the sampling interval, resulting in the lower power limit of the operation of the server. Power Center uses these two values as a power capping range for the server, shown to the user during power cap setup.

Figure 1 shows these key values. The capping value can be set anywhere within the power operation range, but the recommended setting is between the average power consumption and the upper bound value.

Figure 2 shows how these values are displayed in Power Center. Power Center will not allow the capping value to be set outside the lower-bound and upper-bound values.

Figure 1. Power operation range and capping value

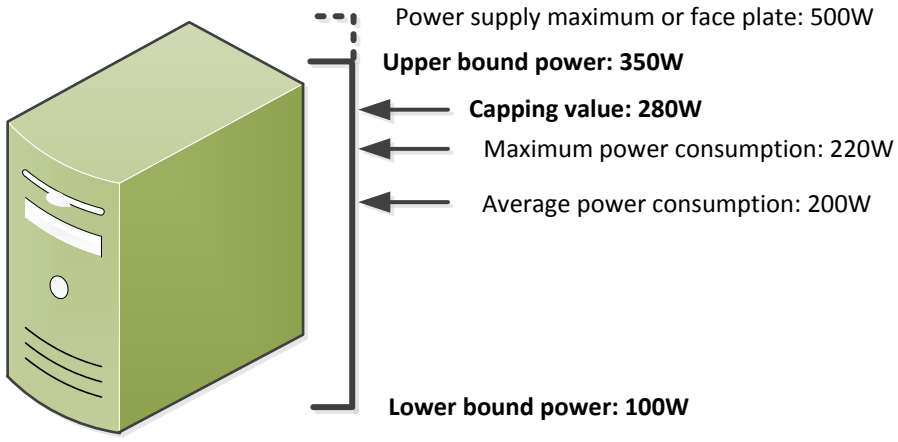
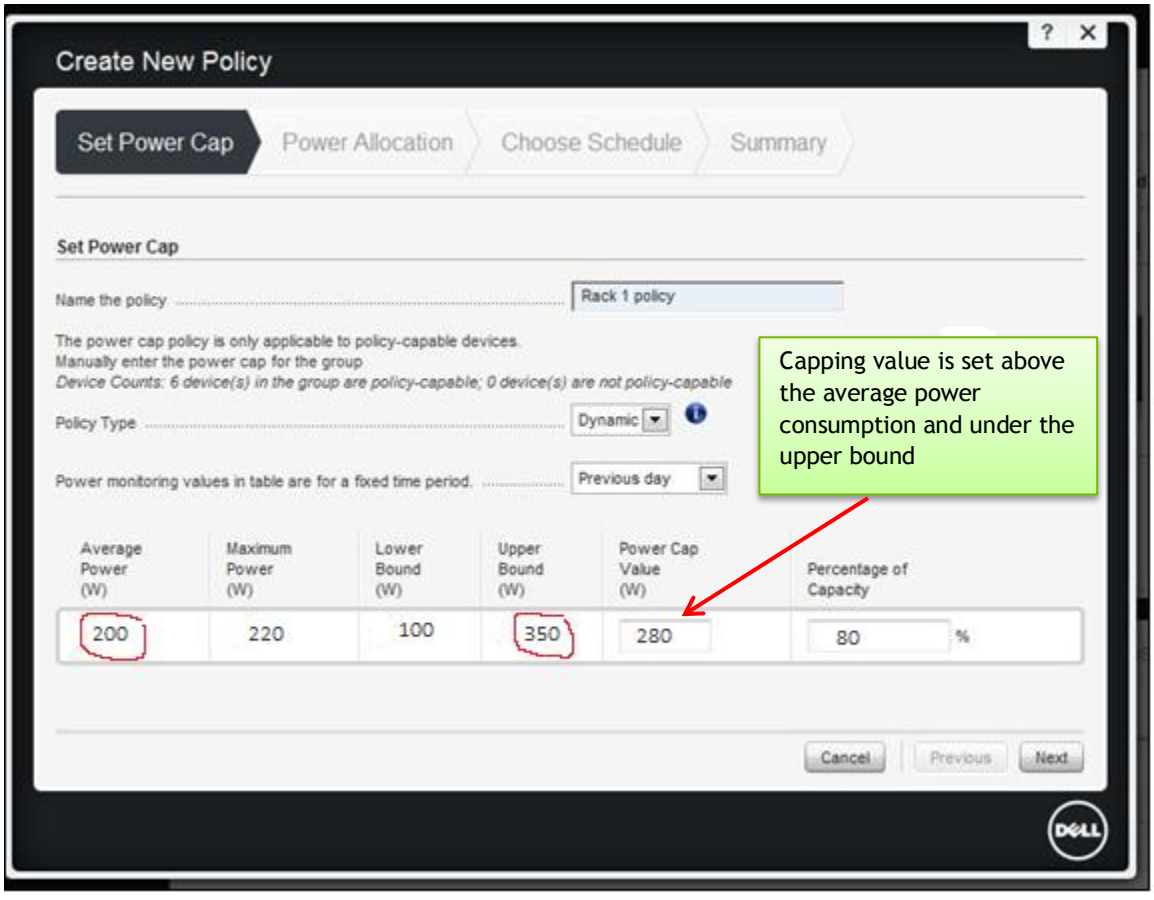

Power supply maximum or face plate: 500W
Upper bound power: 350W
Capping value: 280W
Maximum power consumption: 220W
Average power consumption: 200W
Lower bound power: 100W

Figure 2. Power Center - Power operation range and capping value



Power Capping a Group of Servers

Power Center can Aggregate the Data and Set the Power Capping at any Group Level

Power Center can aggregate power data from servers to provide power monitoring and capping at any group level. It provides both static and dynamic capping for a rack, aisle, room, an entire data center, or any logical group. While power capping helps to limit power usage at any one point in time, it may not reduce an energy bill if it takes longer to process the workload(s) on the throttled server(s); hence, capping should be applied judiciously.

Static Capping

Static capping allows administrators to manually set the capping among the servers in a group. An administrator can set the group capping and decide how to distribute the capping among the servers. The capping value for each server will not be changed until the administrator changes it.

Figure 3 shows a group consisting of three servers. The left configuration shows the range of power operation and capping value of each individual server. The configuration on the right shows how the values are aggregated to the group level.

Figure 3. The power data is aggregated for group capping

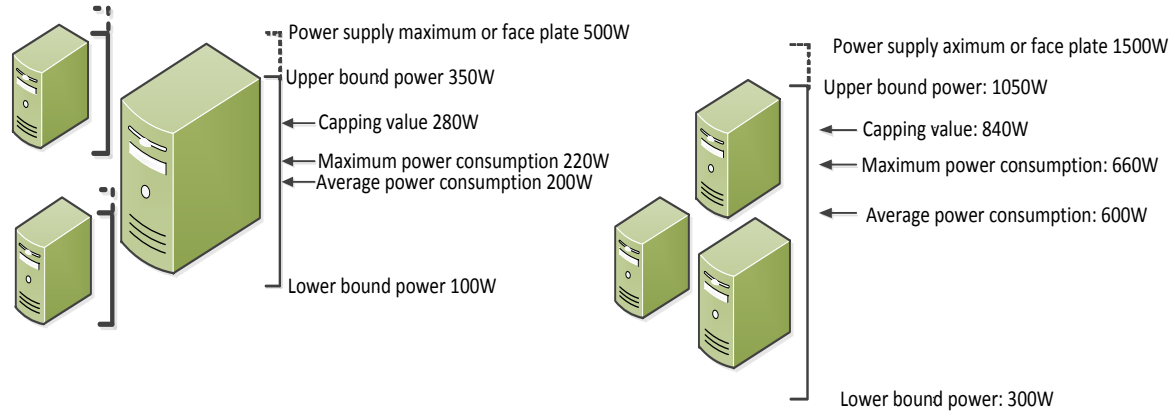
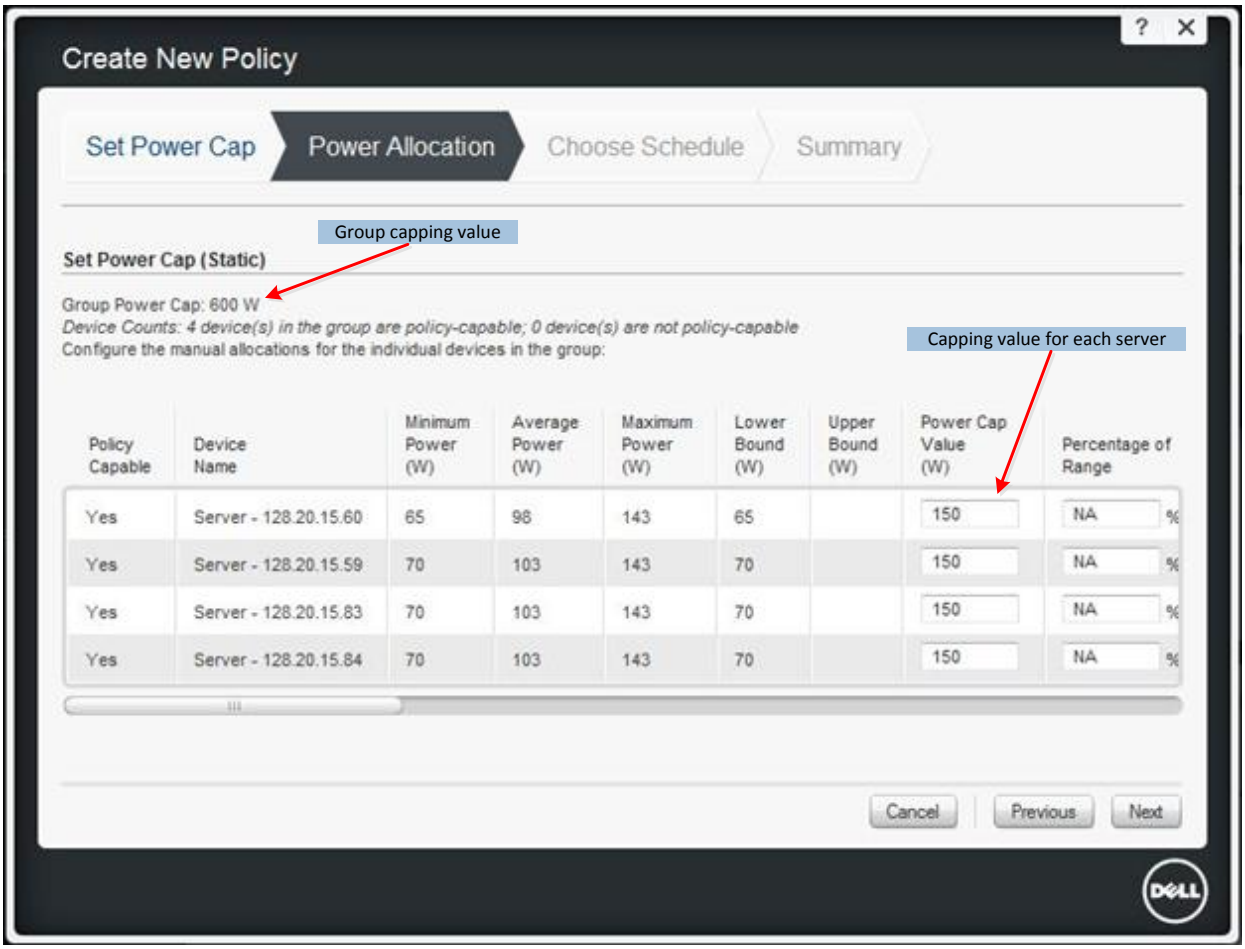

Power supply maximum or face plate 500W
Upper bound power 350W
Capping value 280W
Maximum power consumption 220W
Average power consumption 200W
Lower bound power 100W
Power supply aximum or face plate 1500W
Upper bound power: 1050W
Capping value: 840W
Maximum power consumption: 660W
Average power consumption: 600W
Lower bound power: 300W

Figure 4 shows how static capping is displayed in Power Center. The administrator can set the capping value for each server in the group. The default is an even distribution of power. All power-related data and capping values can be set for several servers at a central console.

Figure 4. Static capping in Power Center



Dynamic Capping

Due to changes in workload(s), the power demand of each individual server can vary greatly and static capping on an individual server basis may not be the best policy as it does not allow a server's power allocation to grow where power is available, resulting in artificial restriction of the system's performance. Unlike static capping, dynamic capping adjusts the capping values of each server in a group, based on the power demand of the server versus the total power allocated and available to the group. Dynamic capping's responsive distribution of power helps to minimize the negative impact to performance.

Dynamic capping is built upon measurement-driven profiling and uses a multi-tiered algorithm to actively adjust the capping values at set intervals. At each sampling interval, Power Center collects the power consumption data from each server. It analyzes the data and, considering the priority setting of the server, adjusts the capping value of each server if power can be reallocated. It raises the power caps of individual servers that are busier and need more power, while lowering the caps for the servers that are not using their total allocated power budget. This balances the group capping and optimizes the use of power among the servers in the group. Note that power is not typically taken from lower-priority servers using their total power budget cap in order to give power to higher priority

servers; however, once excess power is available (known as headroom), the allocation of that power is distributed considering priority.

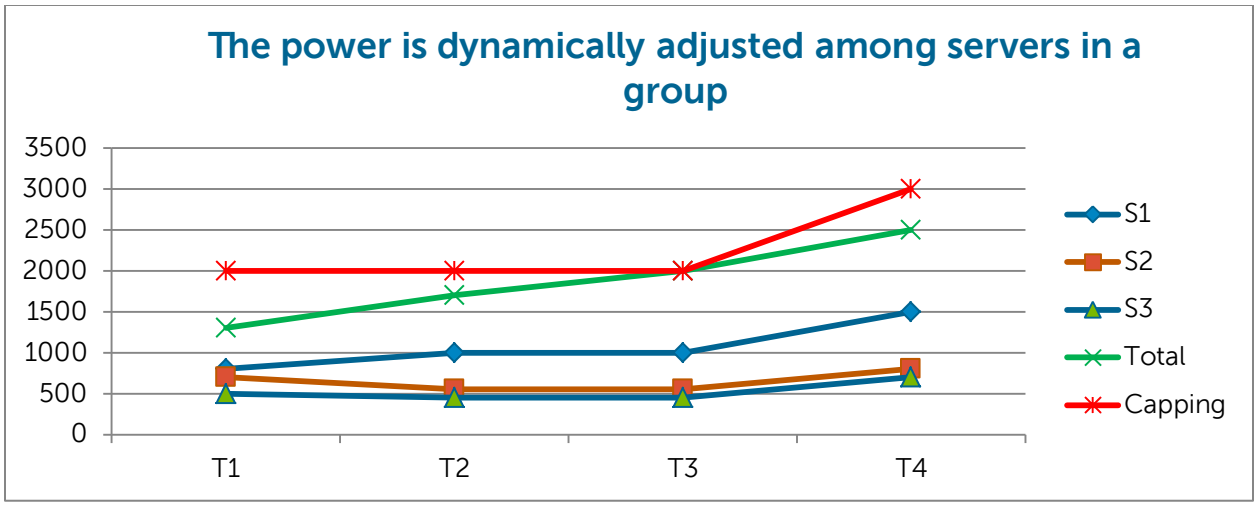
Let's look at an example: Group G1 with 3 servers: S1, S2 and S3, and the group capping is set at 2000W. S1 is set at high priority, while S2 and S3 are set at medium priority. The table below demonstrates how the power capping is adjusted after each sampling interval (T1, T2, etc.). Power capping (P1 column) of each server at each interval is calculated based on the power consumption from the previous interval (P2 column) and is dynamically adjusted as needed. While adjusting the capping value on each server, Power Center also maintains the total power consumption of the group (row G1, column P2) within the group capping ceiling (row G1 and column P1). The worst-case scenario is when the power consumption reaches the budget (interval T3). In this case, an alert is sent, and the administrator will need to take appropriate mitigating action (for example, to increase the group capping to 2500W as in T4 column).

Table 1. The power is dynamically adjusted among servers in a group

	T1		T2		T3		T4		T5	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
S1		500W	800W	800W	1000W	1000W	1000W	1000W	1500W	1200W
S2		500W	700W	500W	550W	550W	550W	550W	800W	700W
S3		300W	500W	400W	450W	450W	450W	450W	700W	600W
G1		1300W	2000W	1700W	2000W	2000W	2000W	2000W	3000W	2500W
E		700W		300W		0				

The table also displays the excess power (row E) or the difference between the power usage and the power cap values. Excess power will be dynamically distributed among the servers in the group based on the priority of each server, with the higher-priority server receiving more power. At interval T1, the excess power of the group is 700W. At the next interval (T2), the power caps are adjusted accordingly. With higher priority, S1 is allocated 300W from the excess, while S2 and S3 each get 200W. At some point during this interval, the power demand in the group increases from 1300W to 1700W, and S1 power consumption reaches its cap. In the next interval (T3), Power Center will take the excess power (from S2 and S3) and redistribute this power to S1. In interval T3, each server has reached its cap limit, and consequently the group has also reached its cap limit. The S1 server will not be allocated more power (even though it has the higher priority) until the cap is increased or the workload in the group is reduced to create some power headroom. The capping value is increased from 2000W to 3000W at the T5 interval. The power is now dynamically redistributed among the servers.

The graph below illustrates how the power is dynamically distributed among the servers in the group and how the total power usage is maintained under the capping value.



How to Set Power Policies using Power Center

The administrator has the option to create multiple capping values and apply different capping values on the server at different times to optimize the use of power resources in the data center. For example, the administrator can create two capping values on a rack, one for the weekdays and another one for the weekend, based on the workload of the rack. Or the administrator may want to create different cap settings for the first shift and second shift workloads. The power policy in Power Center is designed to assist the administrator to combine power capping with scheduling to efficiently distribute power in the data center.

Setting Static Policies

Figures 5 and 6 show the steps to set the static policy using Power Center. The administrator first selects "Static" as the policy type, sets the group capping value, and then manually sets the capping value for each server. Power Center maintains the capping value for each server until the administrator resets the value. The administrator can schedule when the policy takes effect. The administrator can also create multiple policies for a group and apply the policies at different times based on the availability of power resources and the workload of the group.

Figure 5. Setting the static policy type

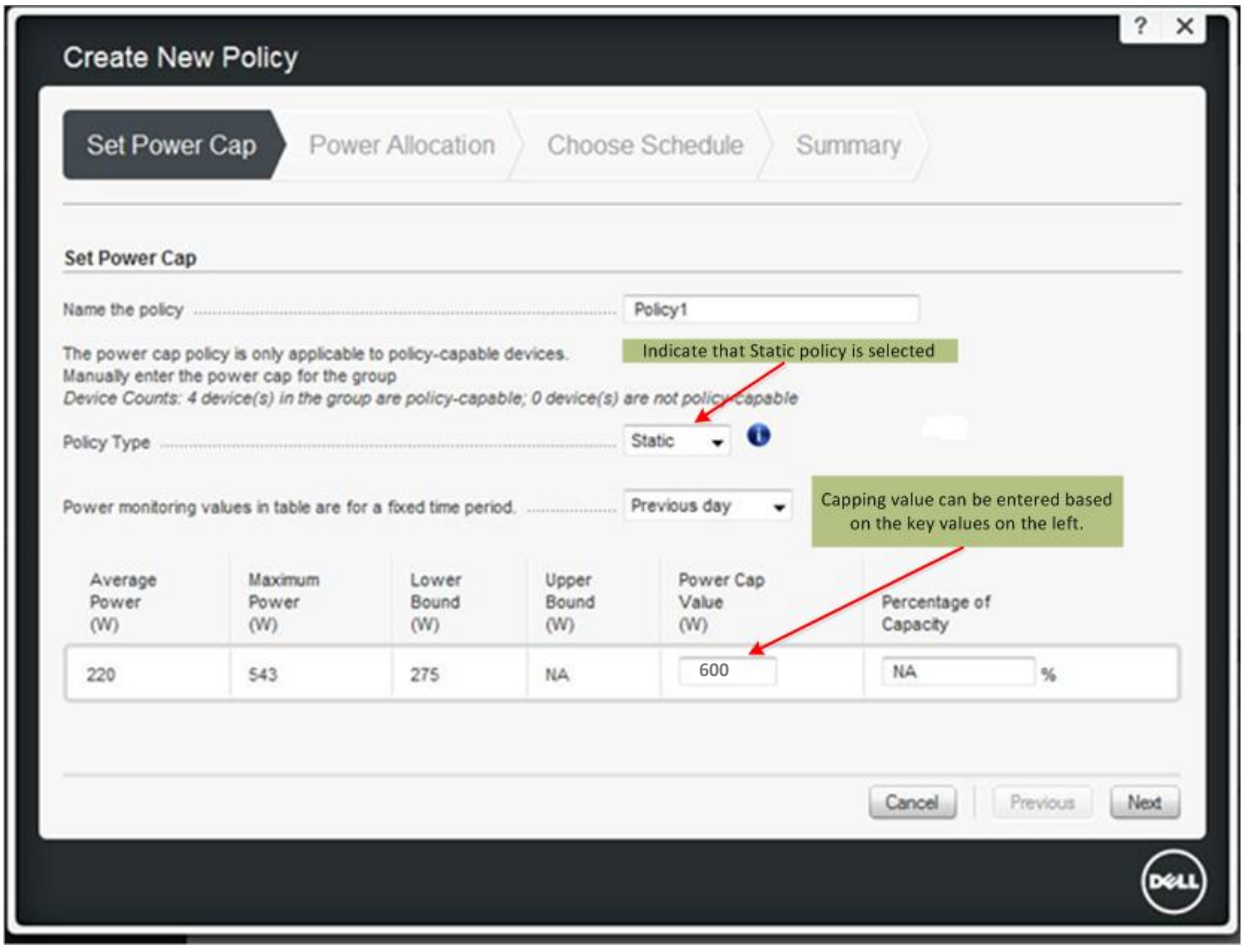
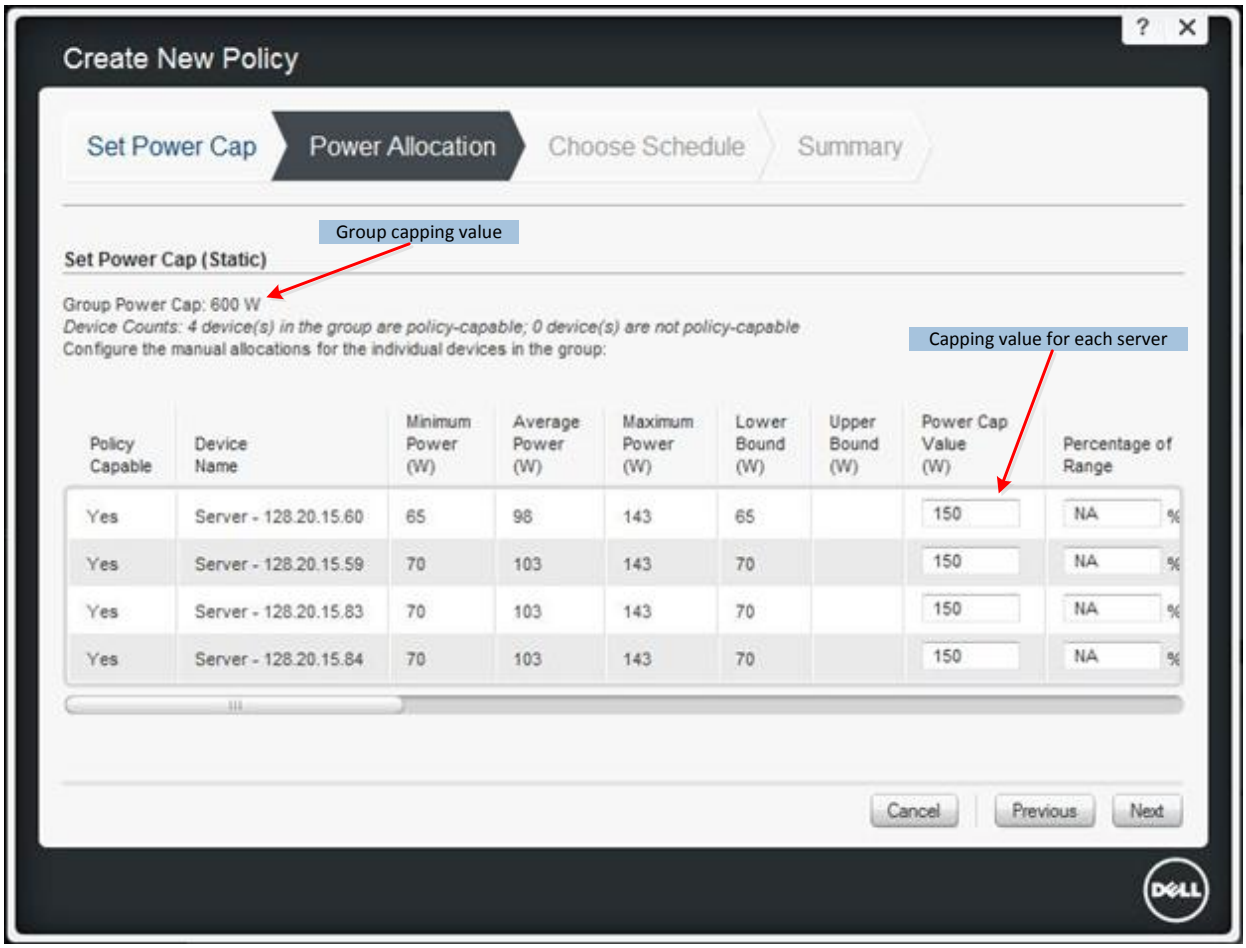


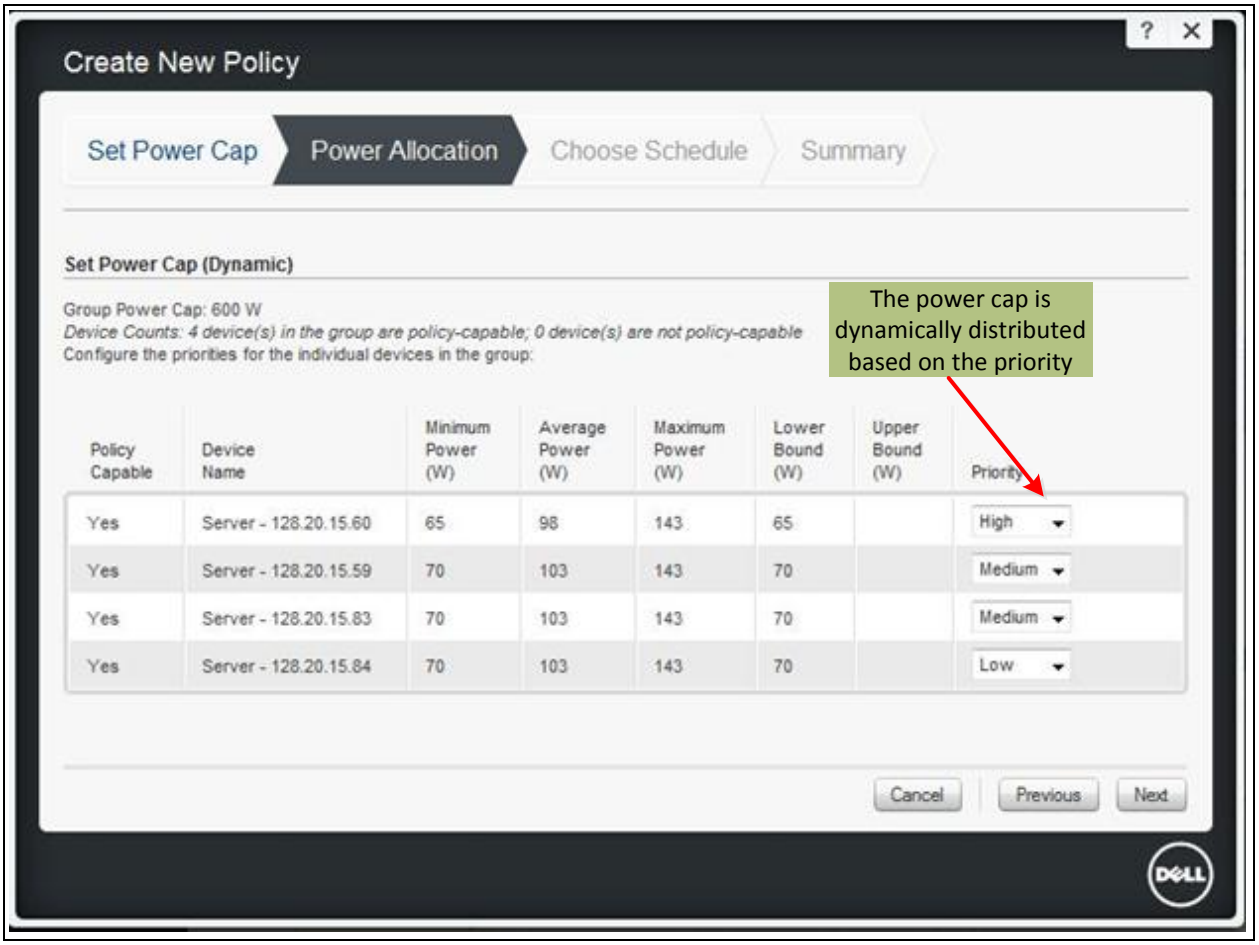
Figure 6. Setting group capping value in static policy



Setting Dynamic Policies

Figure 7 illustrates that when setting the dynamic policy, the user is asked to set the priority for each server, which is different from the static policy, where the user is asked to set the cap value for each server.

Figure 7. Setting priorities in dynamic policy



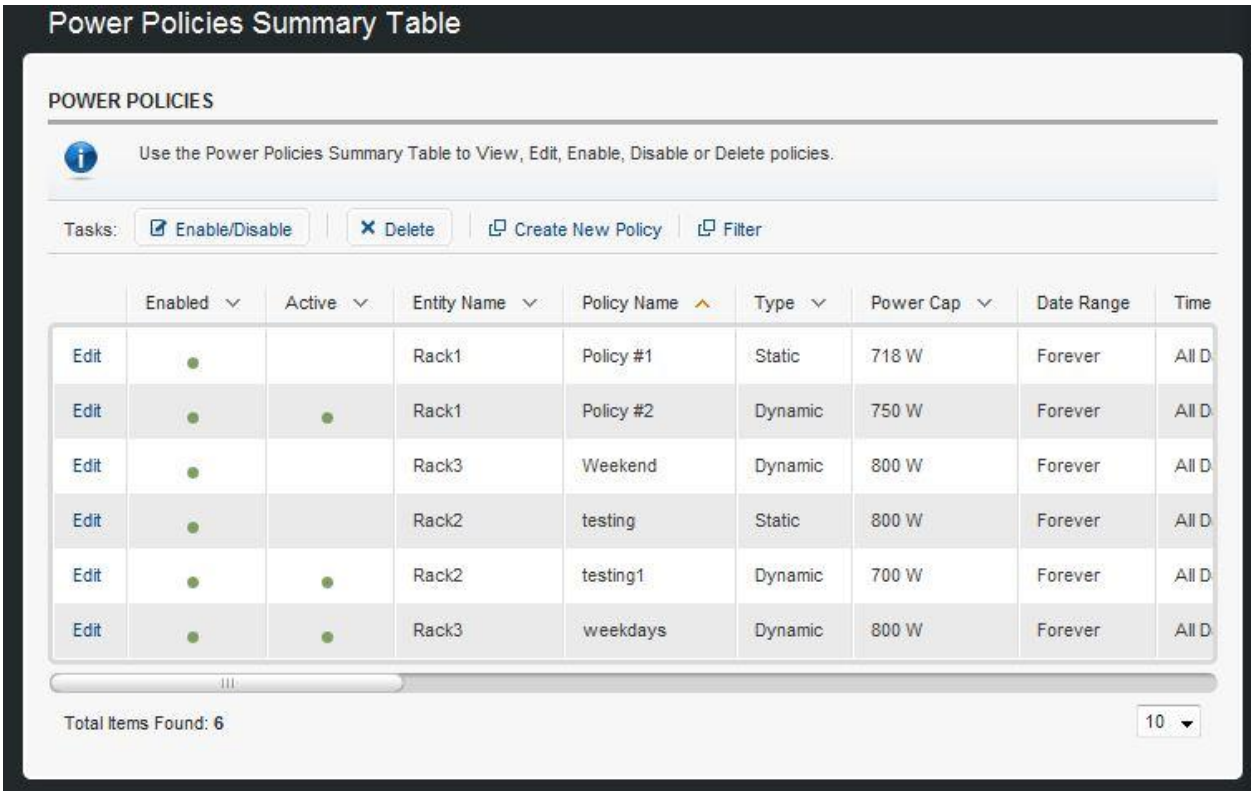
The main purpose of a dynamic policy is to maintain the group budget and to balance the power resources to optimize the performance of the servers in the group, based on recent measurements of power consumption. After measurement of the power consumption of all servers in the group is done, Power Center lowers the cap on servers not using their entire power budget. This creates an excess power budget, known as *headroom*, which is then applied to other servers needing more power in the group, considering priority. After all power caps are calculated, they are sent to Node Manager to be executed on each server. When the total power budget has been allocated and there is no further headroom, a temporary static state is reached where no further power balancing occurs. If a server continues to exceed its power cap, more power will not be allocated to it. Instead, an event is sent to the administrator for appropriate mitigation. The administrator in this case might want to increase the group capping value to allow more power budget for the group or adjust the workloads to keep the servers working under their imposed power caps. Otherwise, keeping the server at an unachievable cap level will impact the performance of the servers, and the power budgeting process may no longer be effective, possibly leading to power circuit breakers being tripped if power is truly constrained.

Managing Power Policies

Power Center provides the aggregated data for power monitoring at the group, rack, row, or room level, or an entire data center. Based on the group hierarchy and power monitoring, users can set the appropriate power policy for each group. However, if there are many groups and the hierarchy becomes overly complex, it would be difficult to track and manage the power policy for each group. To address this problem, Power Center provides a global view for all of the power policies. Figure 8 illustrates how all the power policies are presented in one global table for power policy management in Power Center.

From the table, users can view all of the currently-enabled power policies. For each policy, the table displays the operational status, the group to which the policy applies, the power capping values, and the scheduling information. To change any one of the power policies, users can simply click the "Edit" option on the far left of the policy.

Figure 8. View, manage, and mitigate all power policies from one place


Power Policies Summary Table
POWER POLICIES
Use the Power Policies Summary Table to View, Edit, Enable, Disable or Delete policies.
Tasks: Enable/Disable Delete Create New Policy Filter
Enabled Active Entity Name Policy Name Type Power Cap Date Range Time
Edit Rack1 Policy #1 Static 718 W Forever All D
Edit Rack1 Policy #2 Dynamic 750 W Forever All D
Edit Rack3 Weekend Dynamic 800 W Forever All D
Edit Rack2 testing Static 800 W Forever All D
Edit Rack2 testing1 Dynamic 700 W Forever All D
Edit Rack3 weekdays Dynamic 800 W Forever All D
Total Items Found: 6
10

Power Provisioning Using Real-Time Versus Nameplate Data

A server is normally tagged with a nameplate rating that indicates the maximum power draw of that server. The main purpose of this rating is to inform the administrator of the power infrastructure required to supply power to that server without any consideration of power mitigation. It is typically rated by the manufacturer for the worst-case power draw of all components in a fully-utilized configuration. As such, it is a conservative number that is virtually guaranteed not to be reached.

Hence, power over-provisioning is a common issue in power management when using the nameplate instead of the actual measured power usage values. It is also one of the best and easier ways to first address power utilization in the data center.

Consider the following example: A server has a nameplate capacity at 500W and a total rack capacity of 5000W, allowing the administrator to deploy 10 servers. However, when monitored, Power Center's power consumption graph reports the average power consumption level of each server to be 150W and the maximum power consumption level to be 200W for the duration of a week. Therefore 20 - 25 servers could actually be added to the rack. Compared to the 10 servers, the rack density is increased by 150%. The efficiency rate may be increased even higher if power consumption is monitored and adjusted over time. In many data centers, expansion is limited because the administrator does not have the actual information on how much power their operations are using and how much spare power is available.

Note: The example above is used to illustrate how to address over-provisioning power to a rack. The numbers listed are used as examples only and the numbers will depend on the types of servers.

Note: For the examples above and below, it is recommended to allow some extra (10-20%) power capacity due to unforeseen circumstances. If power capping is expected to be immediately effective when inserting the server in a rack, the server power caps should be set up ahead of time through the iDRAC or through Power Center before insertion into the rack. It is recommended to power-up servers in a rack in a staggered fashion to assure no over-subscription to power during bootup.

Note: For the example below, note that only Dell 12th-generation PowerEdge servers currently support power capping, while Dell 11th- and 12th-generation PowerEdge servers support power monitoring (PMBus power supplies required).

The following is an example of the power deployment steps administrators would take to control and optimize the power consumed by servers in a rack:

1. First, set up a *Rack* group in Power Center and apply a power capping policy of 2000W to that rack. Once servers are set up in Power Center and become a part of the *Rack* group, the power capping policy setup for the rack can help prevent power consumption spikes from tripping the power breaker.
2. Deploy servers to the rack and start monitoring power consumption. Since the cap is set at 2000W, if the total power usage in the rack is above the cap limit, Power Center will attempt to maintain the power usage under 2000W and generate power cap alerts in the console.
3. Analyze the power consumption trends of servers in the rack using Power Center.
4. Mitigate and adjust the power policy as needed using Power Center.

Administrators would repeat the last three steps until the optimal power usage is achieved in the rack. This process can be similarly applied for power deployment to an aisle, room or an entire data center to achieve an optimal power deployment strategy.

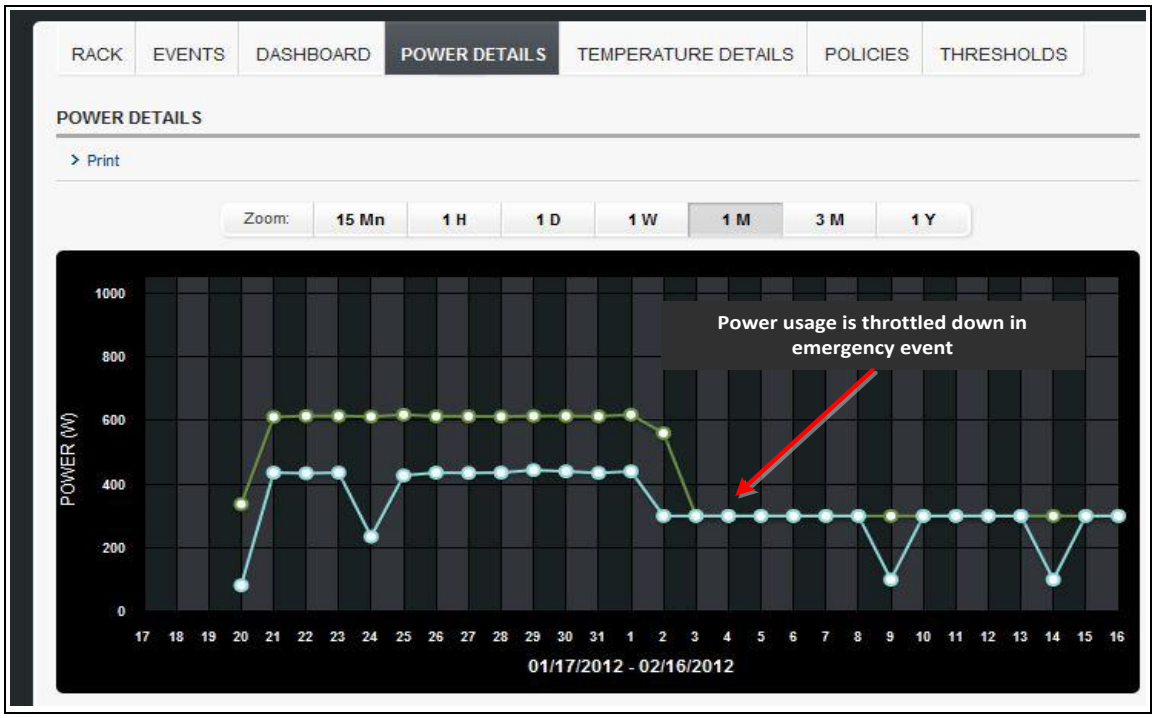
Using Power Policies for Emergency Management

Emergency Power Reduction (EPR) allows the administrator to do the maximum power throttling on a selected number of servers to minimize their power usage in case of an emergency. For example, during a critical event, such as a brownout or a cooling failure, the administrator may want to

minimize the power usage of all the servers in a rack. From the Power Center user interface, the user navigates to the physical-based group of that rack in Power Center and then applies the EPR feature to it. Note that EPR is not meant to be used for long periods of time and certain Power Center configuration features may be disabled during the application of EPR.

Figure 9 shows a rack under EPR. The power usage on the rack's servers is reduced to a minimum.

Figure 9. Power consumption is reduced due to EPR



Setting a Low or Untenable Power Cap on the Server

A power policy is a powerful vehicle for controlling power consumption. The user can virtually set the capping value anywhere between the lower bound and upper bound to restrict the power consumption of a group of servers. However, Dell does not recommend setting the cap value at or below the average power consumption. The average power consumption of a server might rise over time for any number of reasons, including an increase in the workload or increases in fan activity as the temperature of the server rises. When power consumption reaches or rises above the cap value, Power Center will attempt to keep the power consumption at the cap value and send an alert that the power cap is unachievable. Keeping the server at an unachievable cap level will negatively impact the performance of the server and make the power policy ineffective.

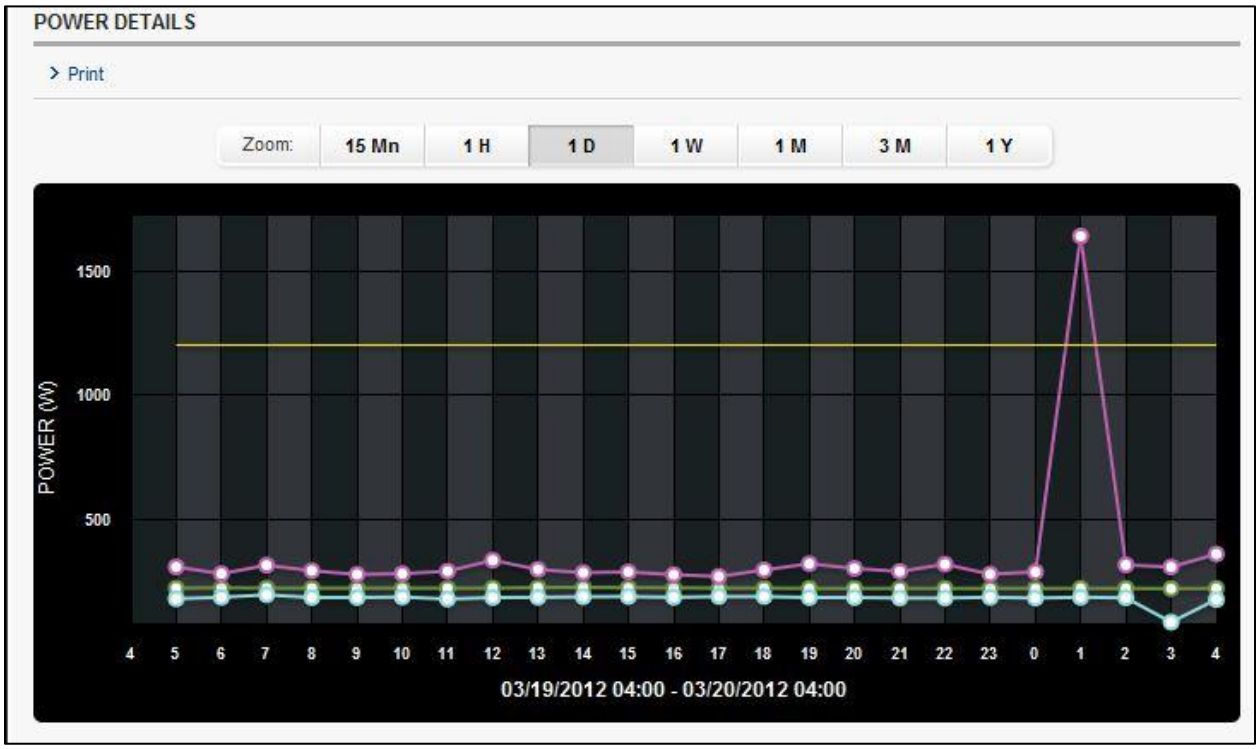
The best practice is to evaluate the trend of power consumption by monitoring the power consumption through Power Center's trend-graphing features, and then to set the cap between the average power consumption and the upper-bound value.

Peak Power Reporting and Power Capping

Power Center measures and reports the power consumption data, including the minimum, maximum and average power consumption of servers, and presents them on graphs at each sampling period; for example, every 15 minutes, 1 hour, etc.

In Figure 10, maximum power consumption (pink line in the graph) is at one point above the power capping value (yellow straight line in the graph). Depending on the situation, there may be a small delay before Node Manager detects the spike, and due to this delay, this spike can sometimes be observed. Generally, the circuit breaker trigger condition is decided by the power accumulated during a short period, not by the maximum power, and Node Manager is designed to throttle the power in a matter of milliseconds. Therefore, the sudden spike on the maximum power usage would likely not trigger the tripping of the circuit breaker as long as the average power usage (the green line in the figure) is maintained under the capping value.

Figure 10. Maximum power consumption spikes above the capping value



Using iDRAC in Conjunction with Power Center

Dell iDRAC7 Enterprise is also a tool for monitoring and setting the power capping for an individual server. If both Power Center and iDRAC are used for setting the power cap, there is the potential for a power capping conflict. Because the same power policy identifier is used between Power Center and the iDRAC, Node Manager will simply enforce the last capping value sent to it. Note that Power Center will not send a capping value to the server if it has not changed from its internal calculations/settings, nor does it know if another entity has overridden its last setting.

Dell recommends that Power Center be exclusively used for setting the power policy for a server if that server has been under Power Center management.

Power Capping Overlap and Logical Groups

Power capping overlap is when multiple power policies are applied to a server due to the groups it is a part of. It is possible for a server to belong to multiple logical groups as well as a physical group; for example, a server can belong to the *Austin* servers group (location) and the *Payroll* servers group (function). When setting a power policy for each group, this can result in a situation where a server has multiple policies and multiple power capping values that could potentially apply at the same time. When there is a power capping overlap between multiple policies in Power Center, the most restrictive power cap will be applied.

Conclusion

Dell OpenManage Power Center enables administrators to see real power usage statistics in their data centers, and to restrict the power usage of Dell 12th-generation PowerEdge servers where needed. Through the ability to group servers and apply a power policy to the group (or just the server itself), the administrator can better manage energy usage to yield cost savings or avoid power interruptions to the data center. Finally, through Power Center's power monitoring and static power policy capabilities, the administrator can plan for a more dense data center, potentially saving space and cooling costs.

References

1. Dell OpenManage Power Center is available for download at http://www.dell.com/powercenter.
2. The User's Guide and other documents related to Dell OpenManage Power Center are available at http://support.dell.com/. Navigate to **Software** -> **Systems Management** -> **Dell OpenManage Power Center**.